Genet. Res., Camb. (2002), **79**, *pp.* 111–118. *With 3 figures.* © 2002 Cambridge University Press
DOI: 10.1017/S0016672302005566 Printed in the United Kingdom

Characterization of the genome of the mealybug *Planococcus lilacinus*, a model organism for studying whole-chromosome imprinting and inactivation

K. NAGA MOHAN[1,2], PARAMITA RAY[1] AND H. SHARAT CHANDRA[1,2]*
[1] *Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560 012, India*
[2] *Centre for Human Genetics, G-5, International Technology Park, Whitefield Road, Bangalore 560 066, India*

(*Received 9 May 2001 and in revised form 9 November 2001*)

Summary

The co-occurrence of three chromosome-wide phenomena – imprinting, facultative heterochromatization and diffuse centromere – in the mealybug *Planococcus lilacinus* makes investigation of the genomics of this species an attractive prospect. In order to estimate the complexity of the genome of this species, 300 random stretches of its DNA, constituting ~ 0·1 % of the genome, were sequenced. Coding sequences appear to constitute ~ 53·5 %, repeat sequences ~ 44·5 % and non-coding single-copy sequences ~ 2 % of the genome. The proportion of repetitive sequences in the mealybug is higher than that in the fruit fly *Drosophila melanogaster* (~ 30 %). The mealybug genome (~ 220 Mb) is about 1·3 times the size of the fly genome (~ 165 Mb) and its GC content (~ 35 %) less than that of the fly genome (~ 40 %). The relative abundance of various dinucleotides, as analysed by the method of Gentles and Karlin, shows that the dinucleotide signatures of the two species are moderately similar and that in the mealybug there is neither over-representation nor under-representation of any dinucleotide.

1. Introduction

In mealybugs and related coccids (Coccoidea; Homoptera; Insecta) there is a system of chromosome behaviour termed 'lecanoid' whose unusual features are of considerable interest (Schrader, 1921; Hughes-Schrader, 1935, 1948; Brown, 1959; reviewed in: Chandra & Brown, 1975, Brown & Chandra, 1977; Nur, 1981). In the majority of mealybug species the diploid number of chromosomes is 10, and there are no sex chromosomes (Brown, 1959, 1961). During early development, male and female embryos are indistinguishable and possess morphologically identical chromosome complements. However, as adults the two sexes show extreme sexual dimorphism: adult males are small and possess a pair of wings whereas females are much larger, wingless and larva-like (Fig. 1). Maleness is associated with inactivation of a haploid set of chromosomes during embryonic development. At around the blastoderm stage of development, 5 of the 10 chromosomes turn heterochromatic and become genetically inactive in those embryos that develop into males. Genetic and cytological studies have demonstrated that in sexually reproducing mealybugs the set of chromosomes which turns heterochromatic and inactive is paternal in origin (Brown & Nelsen-Rees, 1961). These results

* Corresponding author. Tel: +91 80 3601382. Fax: +91 80 3602697. e-mail: sharat@mcbl.iisc.ernet.in

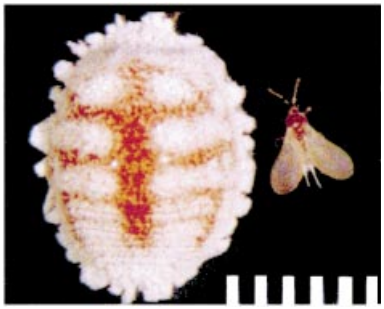

Fig. 1. Sexual dimorphism in *Planococcus lilacinus*. *Left*: Adult female. Since this particular specimen was gravid, it was slightly bigger than virgin females of the same age and from the same culture. *Right*: The male possesses wings, antennae and other features typical of adult insects. (From Kantheti, 1994; original photograph by Dr K. S. Jayarama). Each unit in the scale represents 0·5 mm.

represent one of the earliest demonstrations of genomic imprinting. In interphase nuclei of male insects the paternal set of chromosomes appears as a heterochromatic chromocentre and the 5 maternal chromosomes appear as diffuse euchromatin. Since half of his genome is inactive, the male is said to be a 'physiological haploid' (Hughes-Schrader, 1948). Sex determination is therefore thought to be of the haplo-diploid type (Hughes-Schrader, 1948; Chandra, 1963*b*; Nur, 1963; Chandra & Brown, 1975; Brown & Chandra, 1977). There is no recombination in the male; he transmits only his mother's chromosomes; his father's chromosomes are eliminated during the second meiotic division. There is recombination in the female, and meiosis is inverted, the first being equational and the second reductional (Chandra, 1962).

Experiments involving irradiation of parents prior to mating have shown that paternally derived chromosome fragments, even those that are very small, become heterochromatic in sons (Chandra, 1963*a*), suggesting that genetic elements regulating imprinting and chromosome inactivation are most likely distributed throughout each chromosome. Since radiation-induced chromosome fragments are capable of metaphase alignment and anaphase separation, DNA sequences concerned with centromeric activity also are thought to be distributed throughout each chromosome (Brown & Nelsen-Rees, 1961; Chandra, 1963*a*). There is one other feature of the mealybug genome that is of interest. Mealybugs survive very high doses of ionizing radiation and reproduce thereafter (Chandra, 1963*a*, *b*). They must therefore possess a highly efficient machinery for DNA repair.

Much of our early understanding of genome complexity of model eukaryotes was based on analytical centrifugation of genomic DNA and the kinetics of reassociation of the different fractions (Britten & Kohne, 1968; Britten & Davidson, 1973). Pruitt & Meyerowitz (1986) proposed shotgun sequencing of DNA as an alternative to this method and demonstrated its utility in their study of genome complexity in the mouse-ear cress *Arabidopsis thaliana*. They showed that it was possible to determine the broad features of the *A. thaliana* genome by sequencing as few as 50 randomly chosen recombinant clones (average insert size $\sim$ 12·80 kb; $\sim$ 0·05% of the genome) and that the results so obtained gave a reliable estimate of genome complexity. The complexity of the genome of the puffer fish *Fugu rubripes* was determined by Brenner *et al.* (1993) in a similar manner by analysis of $\sim$ 600 random stretches of genomic DNA from a library containing small inserts (average size $\sim$ 200 bp). By probing the genomic library with single-copy sequences, this method also allowed the estimation of the genome size of *F. rubripes*.

We describe in this report certain features of the mealybug genome as revealed by analysis of 300 genomic shotgun sequences. We compare these features with those of the *Drosophila* genome as ascertained by a similar analysis of 100 random stretches of DNA downloaded from the fly genome database. The sample size upon which this characterization of the mealybug genome is based represents $\sim$ 0·1% of the genome.

2. Methods

(i) *Mealybug shotgun sequences*

Mealybug genomic DNA was digested to completion with *Sau*3A1 and cloned into λ ZAP Express vectors (Stratagene) whose maximum cloning capacity is 10 kb. The *Sau*3A1 fragments ranged in size from $\sim$ 300 bp to $\sim$ 4·0 kb. To estimate the proportion of the genome represented in these fragments, a Southern blot of *Sau*3A1-digested mealybug DNA was probed with a pool of cloned sequences. Most of the hybridization occurred with DNA in the same size-range as that of the *Sau*3A1-digested DNA (data not shown). Little or no hybridization was detected in the size range $>$7·0 kb, suggesting that the mealybug genome was adequately represented in the *Sau*3A1 fragments used to construct the library. This library was excised *in vivo* to obtain a plasmid library in pBK CMV. Three hundred recombinant plasmids were identified from among randomly chosen colonies by digestion with appropriate restriction enzymes. These recombinants were sequenced by both radioactive (manual cycle sequencing) and dye-termination (semi-automated; Perkin Elmer) methods. We refer to these segments of cloned DNA as mealybug shotgun sequences (MSS). Another library was prepared in the same vector with DNA that had been partially digested with *Sau*3A1. The inserts in this library ranged in size from $\sim$ 7·0 kb to $\sim$ 9·0 kb. This library was used for genome size estimation by screening with single-copy sequences identified among the MSS by the method of Brenner *et al.* (1993).

(ii) *Fly shotgun sequences*

The entire sequenced portion of the fly genome (Adams *et al.*, 2000) is represented as 134 assembled segments of DNA termed 'scaffolds' which range in size from $\sim$ 30 kb to $>$ 1 Mb (Myers *et al.*, 2000). All scaffolds in the database were downloaded (www.ncbi.nlm.nih.gov/entrez/query?CMDsearch&DB=genome). From among them, 100 segments (average size $\sim$ 500 bp) were randomly chosen for analysis. The total length of the chosen segments was 50·2 kb, which constitutes $\sim$ 0·03% of the genome. We refer to this subset of sequences as fly shotgun sequences (FSS). The number of FSS selected per chromosome was

https://doi.org/10.1017/S0016672302005566 Published online by Cambridge University Press

proportional to the fraction of the genome represented in that particular chromosome. It was ensured that the FSS were more or less uniformly distributed along each chromosome.

(iii) *Dinucleotide analysis and genome signature*

The proportions of the various dinucleotides were determined by nearest neighbour analysis. Their individual frequencies and relative abundances (ρ^*) were calculated by the method of Gentles & Karlin (2001):

$$\rho^*_{XY} = f_{XY}/f_X f_Y,$$

where f_{XY} is the frequency of the dinucleotide XY, f_X the frequency of the nucleotide X, and f_Y the frequency of nucleotide Y. A ρ^*_{XY} value $\leqslant 0{\cdot}78$ indicates that XY is significantly under-represented and a ρ^*_{XY} value $\geqslant 1{\cdot}23$ suggests that XY is over-represented.

From the values of ρ^*_{XY}, dinucleotide relative abundance distance (δ^*) was calculated between DNA sequences as follows (Gentles & Karlin, 2001):

$$\delta^*(p, q) = \frac{1}{16}\sum_{XY} |\rho^*_{XY}(p) - \rho^*_{XY}(q)| \times 1000,$$

where p and q are two sequences under comparison. The values of δ^* in fly DNA range from 42 to 68 (Gentles & Karlin, 2001). Hence, in the present study involving comparisons of the values of dinucleotide relative abundance (ρ^*) between FSS and other sequences, a δ^* value > 70 was considered significant.

The CpG ratio, which is a measure of the representation of CpG dinucleotides, was determined by the method of Gardiner-Garden & Frommer (1987):

$$\text{CpG ratio} = \frac{\text{Total number of CpG residues} \times \text{Total length of the sequence}}{\text{Total number of Cs} \times \text{Total number of Gs}}.$$

If this ratio is $< 0{\cdot}6$, there is said to be suppression of CpG dinucleotides, and if it is $\geqslant 0{\cdot}6$, there is said to be no CpG suppression.

(iv) *Sequence analysis*

Shotgun sequences were analysed using BLASTX to identify coding sequences with homologues in the SwissProt database. The remaining sequences were classified as non-coding sequences and analysed using the REPEAT program (GCG sequence analysis software) to identify repetitive sequences. The minimum size of the repeat unit was set as 10 bp and no mismatches were allowed. The identified repeating units were classified into microsatellites (sequences with mono-, di-, tri-, …, decanucleotide repeats; Table 1) and minisatellites (sequences with repeating units of complex nature and of length $\geqslant 20$ bp; Table 1). The proportions of microsatellites within exons, introns and intergenic regions were determined following the method of Toth *et al.* (2000). For this purpose, non-coding sequences were considered as intergenic regions. Within coding sequences open reading frames (ORFs) of size $\geqslant 100$ bp were taken as exons and the remainder (i.e. those regions outside ORFs) were taken as introns.

(v) *Southern hybridization*

Genomic DNA samples digested with restriction enzymes were resolved on 1·0% agarose gel, transferred under alkaline conditions by the downward capillary method to Hybond N+ membrane, and cross-linked by using Stratalinker (Stratagene, USA). Prehybridization, hybridization and post-hybridization washes were carried out at 65 °C in high-SDS buffer (7% SDS, 5× SSC, 50 mM sodium phosphate, pH 7·0) following the methods described by Sambrook & Russell (2001).

3. Results

We sequenced 300 clones that were randomly chosen from a genomic library obtained by digesting mealybug DNA with *Sau*3A1. These sequences, which we refer to as mealybug shotgun sequences (MSS), totalled ~210 kb. One hundred randomly downloaded segments of fly DNA (~50 kb), which we refer to as fly shotgun sequences (FSS), were analysed for comparison with MSS. The GC content of MSS is ~35% and that of FSS is ~40%, suggesting that mealybug DNA is more AT-rich than fly DNA. The GC content of MSS reported here is comparable to the values obtained earlier by Deobagkar *et al.* (1982) for *P. lilacinus* geonomic DNA (33.7% by dinucleotide analysis and 32.9%, by reassociation kinetics).

(i) *Dinucleotide composition*

The proportion of dinucleotides was determined by nearest neighbour analysis. Among FSS CpG is the most rare dinucleotide (4·2%), whereas among MSS GpC is the rarest (3·3%). There is no CpG suppression in either MSS (CpG ratio ~1·08) or FSS (CpG ratio ~0·91). TpT/ApA is the most common dinucleotide in both MSS and FSS (20·5% and 16·5% respectively). To examine whether there are differences in dinucleotide composition between FSS and MSS, the relative abundances of the different dinucleotides (ρ^*_{XY}) were determined (see Section 2). As shown in Fig. 2, there

https://doi.org/10.1017/S0016672302005566 Published online by Cambridge University Press

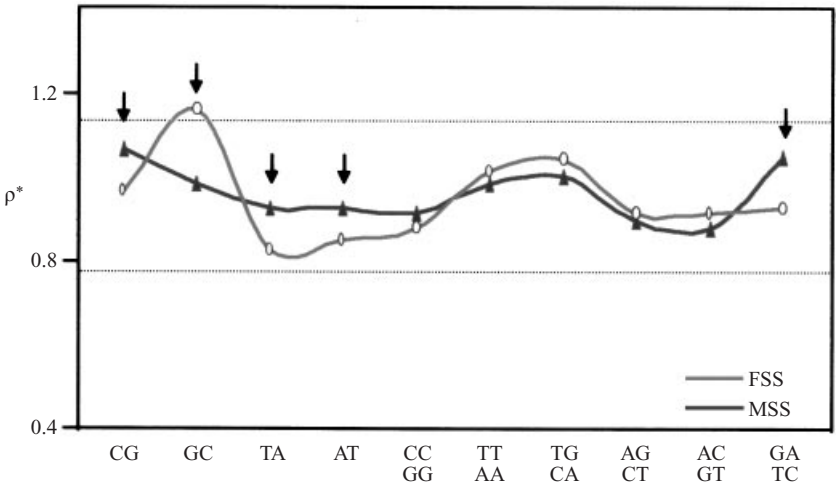

Fig. 2. Analysis of dinucleotides among mealybug shotgun sequences (MSS) and downloaded fly shotgun sequences (FSS). The points (circles and triangles) show dinucleotide relative abundance (ρ*) values in MSS and FSS. Differences in ρ* values between MSS and FSS are higher for the dinucleotides CG, GC, TA, AT and GA/TC (arrows) than for the other dinucleotides.

Table 1. *Repetitive sequences in MSS and FSS*

Nature of the repeat	Repeat unit	MSS		FSS	
		No. of clones	Total length (bp)	No. of sequences	Total length (bp)
Minisatellites	10–15 bp	43	30071	4	1995
	15–20 bp	6	4178	2	1012
	50–100 bp	8	5549	2	1006
	> 100 bp	12	8383	1	489
Microsatellites	$(A–T)_n$, $n \geqslant 10$	42	29645	8	4016
	$(G–C)_n$, $n \geqslant 10$	4	3027	0	
	$(GA–TC)_n$, $n \geqslant 10$	2	1642	5	2502
	$(AGC–GCT)_n$, $n \geqslant 8$	0		3	1489
	$(GAAATTT)_n$, $n \geqslant 7$	1	689	0	
	(A_nT_m), n or $m \geqslant 4$	14	10275	3	1506
Total		132	93459	28	14015

Sequences containing microsatellites constitute ~ 22% of MSS and ~ 19% of FSS. The proportion of sequences containing minisatellites among MSS and FSS is 23% and 10%, respectively.

is neither under-representation nor over-representation of any dinucleotide among MSS (ρ* values range from 0·99 to 1·21). On the other hand, there is significant over-representation of GpC dinucleotides among FSS (ρ* = 1·37). The dinucleotide relative abundance distance (δ*) between FSS and total fly DNA was estimated to be ~ 58·7 (which is not significant); whereas the δ* between FSS and MSS is ~ 73·0. Since δ* within fly sequences varies from 42 to 68 (Gentles & Karlin, 2001), these results suggest that the genome signatures (Karlin & Ladunga, 1994) of the fly and the mealybug are, in the terminology of Gentles and Karlin, 'moderately similar'.

(ii) *Microsatellites and minisatellites*

Twenty-eight of the 100 FSS studied contain repetitive sequences, of which 19 contain microsatellites and 9 contain minisatellites (Table 1). The total length of the microsatellites in FSS is 269 bp (Table 2). As mentioned above, the size of the fly DNA sample studied was ~ 50 kb. On this basis, each megabase of fly DNA is estimated to contain microsatellites of size amounting to 5380 bp (269 × 1000 kb/50 kb). The microsatellites in FSS are distributed more or less equally among coding and non-coding sequences. However, within coding sequences the microsatellites are ~ 2·5 times more frequent in introns than in exons.

One hundred and thirty-two of the 300 MSS studied contain repeat sequences. These repeats correspond to 93459 bp and constitute ~ 44·5% of the MSS. Among the 132 repeat sequences, 63 contain microsatellites (45278 bp, constituting 21·6% of MSS) and the remaining 69 contain minisatellites (48181 bp, constituting 22·9% of MSS). The total length of micro-

https://doi.org/10.1017/S0016672302005566 Published online by Cambridge University Press

Table 2. *Distribution of microsatellites among MSS and FSS*

	MSS				FSS			
	Coding				Coding			
Repeat unit	Exons	Introns	Non-coding	Total	Exons	Introns	Non-coding	Total
Mono	0	19	37	56	0	17	26	43
Di	0	31	42	73	0	29	40	69
Tri	29	14	21	64	27	13	15	55
Tetra	0	12	13	25	0	12	14	26
Penta	0	13	17	30	0	13	15	28
Hexa	12	15	23	50	12	15	21	48
Total (in bp)	41	104	153	298	39	95	135	269

Microsatellites $\geqslant$ 12 bp in length which contain repeating units ranging from 1 bp to 6 bp. In both MSS and FSS the microsatellites occur at more or less the same frequency in coding and non-coding sequences. However, within coding sequences they occur $\sim$ 2·5 times more often in introns than in exons. The total length of microsatellites per 50 kb of MSS is about 298 bp and the corresponding value for the FSS studied is 269 bp.

satellites in MSS is 1252 bp and their estimated length per Mb of mealybug DNA is $\sim$ 5960 bp, about 1·1 times higher than in FSS. This value and the between-species ratio of the proportions of microsatellites (21·6% (MSS)/ 19% (FSS) = 1·1) are identical. The distribution of microsatellites among coding and non-coding sequences within MSS is similar to that in FSS.

The proportion of minisatellites is higher in MSS ($\sim$ 23%) than in FSS ($\sim$ 10%). The minisatellites among MSS do not show homology to any dispersed category of repeat sequences in the databases, suggesting that they may be clustered repeats. The most common minisatellite (6% of the repeat sequences) is 180 bp long and is represented in eight clones. Southern analysis of mealybug DNA with this sequence confirmed that it is a clustered repeat (Fig. 3*A*). A 300 bp repeat was identified in two clones. This sequence contains several internal tandem repeats which range in size from 20 bp to 40 bp. As shown in Fig. 3*B*, this sequence hybridizes to mealybug DNA over a wider size range than does the 180 bp repeat. There are seven clones containing sequences that are both non-repetitive and non-coding. These unique single-copy sequences constitute $\sim$ 2·03% of the MSS. Such single-copy sequences were not found among the FSS.

(iii) *Poly(A/T) stretches and GAAAA repeats*

Forty of the 300 MSS were found to contain $(A/T)_4$ tracts and GAAAA sequences. Of these 40, 16 are non-coding and 24 are coding. Within coding sequences, these features occur twice as often in introns as in exons.

(iv) *Genome size estimation*

One hundred and sixty-one of the 300 MSS studied are coding. The total length of these sequences is 112·34 kb and they constitute 53·5% of MSS. In the case of FSS, 72 of the 100 sequences (36·20 kb), constituting 71·9% of the downloaded sequences, are coding. These results suggest that the proportion of coding sequences in the genome of the mealybug is smaller than in the fly. However, since mealybug and fly are phylogenetically related organisms belonging to the same taxonomic class, it is not unreasonable to expect that the repertoire of genes in the two species would be similar. Since the proportion of coding sequences in the mealybug genome appears to be smaller than that in the fly genome, the size of the mealybug genome is expected to be correspondingly larger (71·9/53·5), i.e. 1·34 times the fly genome. On this basis, we estimate the size of the mealybug genome to be about 221 Mb (1·34 × 165, the size of the fly genome). Genome size was also estimated by screening a mealybug genomic library with 10 single-copy sequences obtained by the shotgun method. The library contained inserts ranging from $\sim$ 7·0 kb to $\sim$ 9·0 kb (see Section 2). The results of this experiment suggest that on an average 25186 plaque forming units represent the size of the mealybug genome. This gives an estimate of $\sim$ 201 Mb for the size of the mealybug genome (25186 × 8·0 kb, the average insert size). Thus, the estimates obtained by the two methods are not dissimilar and are in agreement with that obtained by Muralidharan (1985) by photometry of Feulgen-stained nuclei ($\sim$ 220 Mb).

4. Discussion

The studies of Pruitt & Meyerowitz (1986) on *A. thaliana* and those of Brenner *et al.* (1993) on *F. rubripes* have shown that by sequencing and analysis of random stretches of DNA constituting as little as 0·05% of the genome, it is possible to estimate the broad features of a complex genome. In the present

https://doi.org/10.1017/S0016672302005566 Published online by Cambridge University Press

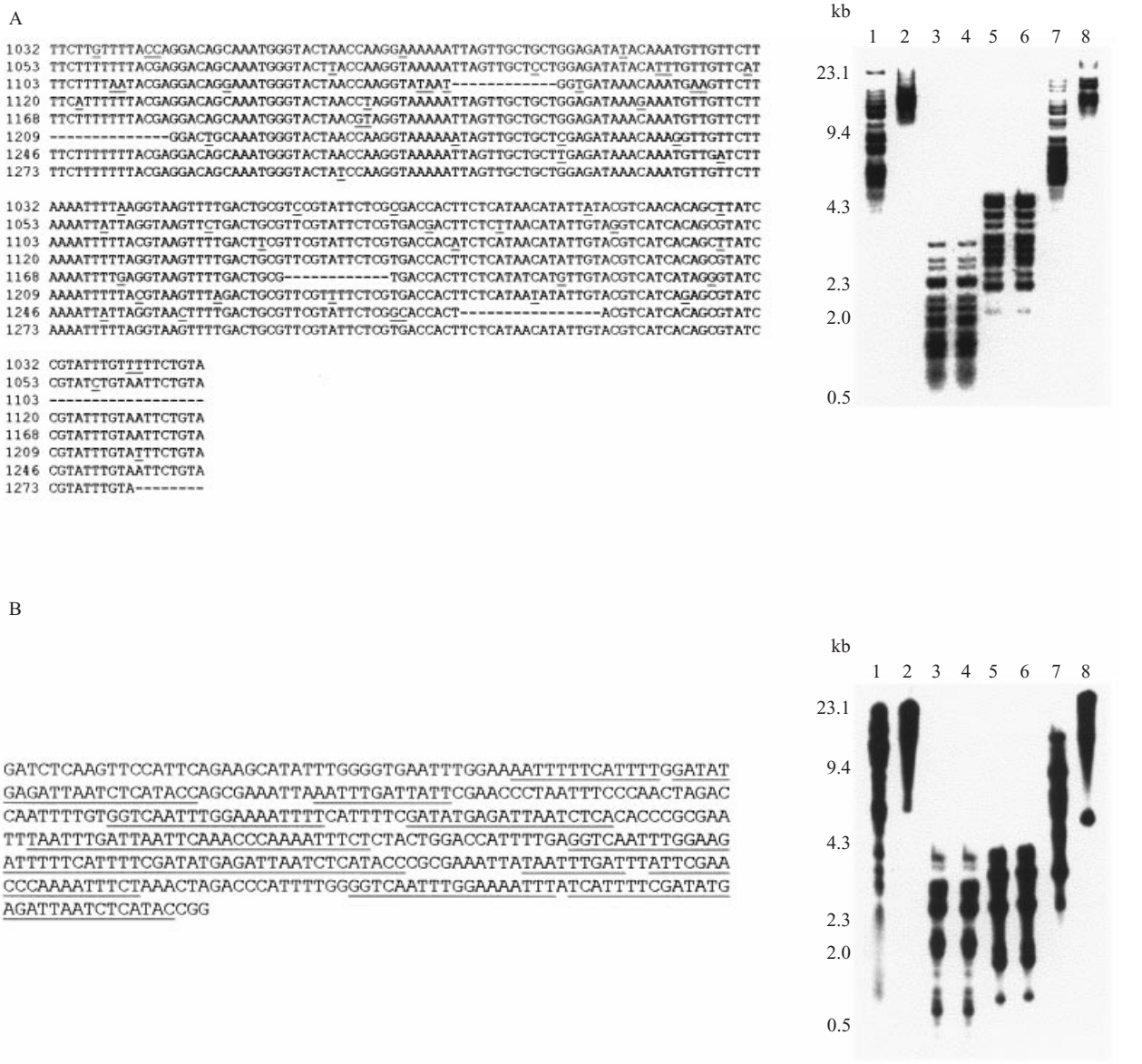

Fig. 3. Two mealybug minisatellites. (*A*) The most common minisatellite among MSS has a repeating unit of 180 bp, and it is present in eight clones (indicated by clone numbers at extreme left). It constitutes 6% of the total repeat sequences in MSS. (*B*) A complex 300 bp repeat. This sequence, represented in two clones, contains several internal tandem repeats (underlined). These internal repeats were also observed in 12 other clones as part of other sequences. The panels at right show hybridization of Southern blots of mealybug DNA, digested with different restriction enzymes (lanes: 1, *Ase*I; 2, *Hin*dIII; 3, *Hpa*II; 4, *Msp*I; 5, *Sau*3A1; 6, *Mbo*I; 7, *Eco*RI; 8, *Xho*I) to sequences in (*A*) and (*B*). The 300 bp repeat, hybridizes over a wider size-range of DNA fragments than the 180 bp repeat.

study, stretches of fly sequences were randomly chosen from the fly genome database and downloaded. These sequences, constituting ~0·03% of the fly genome, were analysed for determining the relative proportions of coding and non-coding sequences, GC content, dinucleotide composition, proportions of microsatellites and minisatellites, and distribution of microsatellites in coding and non-coding sequences. The estimates obtained from this sample of the fly sequence are nearly identical to those obtained by analysis of the total fly sequence (Kram *et al*., 1971*a*, *b*; Lewin, 1997; Toth *et al*., 2000; Gentles & Karlin, 2001), suggesting that sampling even 0·03% of the genome may allow a reliable estimation of genome complexity. It may be noted that in the present study sequences constituting ~0·1% of the mealybug genome were analysed.

Particular dinucleotides are either under-represented or over-represented in the genomes of multicellular organisms (Gentles & Karlin, 2001). No such bias was found in MSS. A similar lack of bias has been observed in the genome of *Sachharomyces cerevisiae* (Gentles & Karlin, 2001). CpG dinucleotides occur less often in MSS (3·7%) than in FSS (4·2%). However, the GC content of fly DNA (40%) is ~1·14 times that of mealybug DNA (35%) and, as a result,

https://doi.org/10.1017/S0016672302005566 Published online by Cambridge University Press

the proportions of CpG dinucleotides in the two species are similar ($1{\cdot}14 \times 3{\cdot}7 = 4{\cdot}2$). On the other hand, the proportion of GpC dinucleotides in MSS (3·3%) is less ($1{\cdot}14 \times 3{\cdot}3 = 3{\cdot}76\%$) than in FSS (5·4%), in agreement with the findings of Gentles & Karlin (2001) that over-representation of GpC dinucleotides is a feature of fly DNA.

Dinucleotide relative abundance distance (δ^*) is thought to provide information on the degree of similarity between the composition of dinucleotides in two genomes (Gentles & Karlin, 2001). In the present study, δ^* values were calculated between mealybug shotgun sequences (MSS), fly shotgun sequences (FSS) and total fly sequence. δ^* between FSS and MSS was found to be 73·0, $\sim$ 1·25 times the distance between FSS and the total fly sequence (58·7). Since δ^* within fly DNA ranges from 42 to 68 (Gentles & Karlin, 2001), the signatures of the fly and mealybug genomes are, in the terminology of Gentles & Karlin, moderately similar.

The inactive X chromosome in female mammals and the inactive paternal set of chromosomes in coccids are well-known examples of facultative chromosome inactivation. Repetitive sequences have been implicated in such inactivation (Chandra, 1963*a*; Brown & Chandra, 1973; Gartler & Riggs, 1983; Riggs, 1990; Lyon, 1998). In mammals, inactivation of the X chromosome is regulated by a single locus, *Xist* (X inactivation-specific transcript) (Brown *et al*., 1991; Penny *et al*., 1996) located in the X inactivation centre (XIC) (Russell, 1963; Cattanach, 1974, 1975). Riggs (1990) suggested that repetitive elements serving as 'way stations' or 'boosters' may facilitate the spread of inactivation. Lyon (1998) has proposed that long interspersed nuclear elements (LINEs) may mediate the spreading of inactivation from the XIC. LINEs occur in mouse and human X chromosomes at frequencies twofold higher than expected (Bailey *et al*., 2000) and they are particularly frequent in regions of the X chromosome that undergo inactivation, suggesting a possible role for LINEs in the spread of X inactivation.

Repetitive elements have also been implicated in the imprinting and inactivation of whole chromosomes in mealybugs (Chandra, 1963*a*). The results reported here, that the mealybug genome contains a higher proportion of minisatellites than the fly genome, as well as the earlier results of Kantheti (1994) and Khosla *et al*. (1996, 1999), appear relevant to this hypothesis. Kantheti (1994) observed that $\sim$ 10% of male chromatin is resistant to digestion by micrococcal nuclease. No such nuclease-resistant chromatin (NRC) fraction was observed in females. It was therefore inferred that the most likely source of NRC is the inactive and heterochromatic set of paternal chromosomes. A subsequent study of 16 NRC-associated sequences led to the identification of two repeat sequences: NRC50 and NRC51 (Khosla *et al*., 1999). NRC50 is organized as nuclease-resistant chromatin in males but not in females, whereas NRC51 is nuclease-resistant in both females and males. Further studies on NRC-associated and other repetitive sequences are likely to be of interest in relation to the hypothesis that repetitive sequences are involved in the chromosome-wide phenomena characteristic of the mealybug genetic system.

We thank Mustafa Saifi and Prameela Kantheti for discussions and N. V. Joshi and two anonymous referees for several useful suggestions. The editorial assistance of N. A. Prakash and the assistance of Yogaraja Gowda, K. Srinivasan and K. Kuppuswamy in the laboratory are very much appreciated. DNA sequencing was done at the DBT-funded central DNA sequencing facility (Prof. M. R. S. Rao), Indian Institute of Science. This work was supported by a grant from the Council of Scientific and Industrial Research, New Delhi. K. N. M. received fellowship support from the Centre for Human Genetics during part of this investigation.

References

Adams, M. D., Celniker, S. E., Holt, R. A., Gocayne, J. D., Amanitides, P. G., Scherer, S. E., Li, W. P. *et al*. (2000). The genome sequence of *D. melanogaster*. *Science* **287**, 2185–2195.

Bailey, J. A., Carrel, L. & Chakravarti, A. (2000). Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proceedings of the National Academy of Sciences of the USA* **97**, 6634–6639.

Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B. & Aparicio, S. (1993). Characterization of the *Fugu* (puffer fish) genome as a compact model vertebrate genome. *Nature* **366**, 265–268.

Britten, R. J. & Davidson, E. H. (1973). Organization, transcription and regulation in the animal genome. *Quarterly Review of Biology* **48**, 565–613.

Britten, R. J. & Kohne, D. E. (1968). Repeated sequences in DNA. *Science* **161**, 529–540.

Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R. & Willard, H. F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38–44.

Brown, S. W. (1959). Lecanoid chromosome behavior in three more families of the Coccoidea (Homoptera). *Chromosoma* **10**, 278–300.

Brown, S. W. (1961). Fracture and fusion of coccid chromosomes. *Nature* **191**, 1419–1420.

Brown, S. W. & Chandra, H. S. (1973). Inactivation system of the mammalian X chromosome. *Proceedings of the National Academy of Sciences of the USA* **70**, 195–199.

Brown, S. W. & Chandra, H. S. (1977). Chromosome imprinting and the differential regulation of homologous chromosomes. In *Cell Biology: A Comprehensive Treatise*, Vol. I (ed. L. Goldstein & D. M. Prescott), pp. 108–189. New York: Academic Press.

Brown, S. W. & Nelsen-Rees, W. A. (1961). Radiation analysis of a lecanoid genetic system. *Genetics* **42**, 510–523.

Cattanach, B. M. (1974). Position effect variegation in the mouse. *Genetical Research* **23**, 291–306.

Cattanach, B. M. (1975). Control of chromosome inactivation. *Annual Review of Genetics* **9**, 1–18.
Chandra, H. S. (1962). Inverse meiosis in triploid females of the mealybug, *Planococcus citri*. *Genetics* **47**, 1441–1454.
Chandra, H. S. (1963*a*). Cytogenetic studies following high dosage paternal irradiation in the mealybug *Planococcus citri*. I. Cytology of X_1 embryos. *Chromosoma* **14**, 310–329.
Chandra, H. S. (1963*b*). Cytogenetic studies following high dosage paternal irradiation in the mealybug *Planococcus citri*. II. Cytology of X_1 females and the problem of lecanoid sex determination. *Chromosoma* **14**, 330–346.
Chandra, H. S. & Brown, S. W. (1975). Chromosome imprinting and the mammalian X chromosome. *Nature* **253**, 165–168.
Deobagkar, D. N., Muralidharan, K., Devare, S. G., Kalghatgi, K. K. & Chandra, H. S. (1982). The mealybug chromosome system I: Unusual methylated bases and dinucleotides in DNA of a *Planococcus* species. *Journal of Biosiences* **4**, 513–526.
Gardiner-Garden, M. & Frommer, M. (1987). CpG islands in vertebrate genomes. *Journal of Molecular Biology* **196**, 261–282.
Gartler, S. M. & Riggs, A. D. (1983). Mammalian X-chromosome inactivation. *Annual Review of Genetics* **17**, 155–190.
Gentles, A. J. & Karlin, S. (2001). Genome-scale compositional comparisons in eukaryotes. *Genome Research* **11**, 540–546.
Hughes-Schrader, S. (1935). The chromosome cycle of *Phenococcus* (Coccidae). *Biological Bulletin* (Woods Hole, Mass) **69**, 462–468.
Hughes-Schrader, S. (1948). Cytology of coccids (Coccoidea–Homoptera). *Advances in Genetics* **2**, 127–203.
Kantheti, P. (1994). Studies on a female-specific cDNA clone, and chromatin organization, in a mealybug *Planococcus lilacinus*. PhD thesis, Indian Institute of Science, Bangalore, India.
Karlin, S. & Ladunga, I. (1994). Comparisons of eukaryotic genomic sequences. *Proceedings of the National Academy of Sciences of the USA* **91**, 12832–12836.
Khosla, S., Kantheti, P., Brahmachari, V. & Chandra, H. S. (1996). A male-specific nuclease-resistant chromatin fraction in the mealybug *Planococcus lilacinus*. *Chromosoma* **104**, 386–392.
Khosla, S., Augustus, M. & Brahmachari, V. (1999). Sex-specific organization of middle repetitive DNA sequences in the mealybug *Planococcus lilacinus*. *Nucleic Acids Research* **27**, 3745–3751.
Kram, R., Botchan, M. & Hearst, J. E. (1971*a*). The occurrence of highly reiterated DNA sequences in the genome of *Drosophila melanogaster*. I. Isolation. *Archives Internationales de Physiologie*, *de Biochimie et de Biophysique* **79**, 417–419.
Kram, R., Botchan, M. & Hearst, J. E. (1971*b*). The occurrence of highly reiterated DNA sequences in the genome of *Drosophila melanogaster*. II. Kinetics of reassociation. *Archives Internationales de Physiologiè*, *de Biochimie et de Biophysique* **79**, 419–420.
Lewin, B. (1997). Genomes. In *Genes VI* (ed. B. Lewin), pp. 645–662. New York: Oxford University Press.
Lyon, M. F. (1998). X-chromosome inactivation: a repeat hypothesis. *Cytogenetics and Cell Genetics* **80**, 133–137.
Muralidharan, K. (1985). A study of DNA methylation in the mealybug *Planococcus lilacinus*. PhD thesis, Indian Institute of Science, Bangalore, India.
Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., Kravitz, S. A. *et al*. (2000). A whole-genome assembly of *Drosophila*. *Science* **287**, 2196–2204.
Nur, U. (1963). Meiotic parthenogenesis and heterochromatization in a soft scale, *Pulvinaria hydrangeae* (Coccoidea: Homoptera). *Chromosoma* **19**, 439–448.
Nur, U. (1981). Evolution of unusual chromosome systems in scale insects (Coccoidea: Homoptera). In *Insect Cytogenetics* (ed. R. L. Blackburn, G. M. Hewitt & M. Ashburner), pp. 97–117. Oxford: Blackwell Scientific.
Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. (1996). Requirement for *Xist* in X chromosome inactivation. *Nature* **379**, 131–137.
Pruitt, R. E. & Meyerowitz, E. M. (1986). Characterization of the genome of *Arabidopsis thaliana*. *Journal of Molecular Biology* **187**, 169–183.
Riggs, A. D. (1990). Marsupials and mechanisms of X chromosome inactivation. *Australian Journal of Zoology* **37**, 419–441.
Russell, L. B. (1963). Mammalian X-chromosome action: inactivation limited in spread and in region of origin. *Science* **140**, 976–978.
Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*. 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Press.
Schrader, F. (1921). The chromosomes of *Pseudococcus nipae*. *Biological Bulletin* (Woods Hole, Mass) **40**, 259–270.
Toth, G., Gaspari, Z. & Jurka, J. (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research* **10**, 967–981.